RNA-interference and Register Machines (extended abstract)

Masahiro Hamano
PRESTO , Japan Science and Technology Agency (JST)
4-1-8 Honcho Kawaguchi, Saitama 332-0012, JAPAN.
hamano@is.s.u-tokyo.ac.jp

RNA interference (RNAi) is a mechanism whereby small RNAs (siRNAs) directly control gene expression without assistance from proteins. This mechanism consists of interactions between RNAs and small RNAs both of which may be single or double stranded. The *target* of the mechanism is mRNA to be degraded or aberrated, while the *initiator* is double stranded RNA (dsRNA) to be cleaved into siRNAs. Observing the digital nature of RNAi, we represent RNAi as a Minsky register machine such that (i) The two registers hold single and double stranded RNAs respectively, and (ii) Machine's instructions are interpreted by interactions of enzyme (Dicer), siRNA (with RISC complex) and polymerization (RdRp) to the appropriate registers. Interpreting RNAi as a computational structure, we can investigate the computational meaning of RNAi, especially its complexity. Initially, the machine is configured as a Chemical Ground Form (CGF), which generates incorrect jumps. To remedy this problem, the system is remodeled as recursive RNAi, in which siRNA targets not only mRNA but also the machine instructional analogues of Dicer and RISC. Finally, probabilistic termination is investigated in the recursive RNAi system.

1 Introduction

RNA interference (RNAi), also known as RNA silencing, is a mechanism whereby a small interfering RNA (siRNA) originating from double stranded RNA (dsRNA) directly controls gene expression of a target mRNA [1, 5]. The two key steps of RNAi are:
(i) dsRNA is cleaved into small siRNA's fragments by an enzyme known as Dicer.
(ii) A single strand of one small siRNA is recruited by the argonaute protein to form a complex called RISC. Using the siRNA as a template, RISC then identifies matching sequences in a target mRNA, and induces the mRNA to degrade or become aberrant (see the right semicircle of Figure 1).
Therefore, we can regard the initiator of RNAi as dsRNA (since it supplies the siRNAs) and the target as mRNA (to be degraded or aberrated by a siRNA in a Watson-Crick complementary manner).

A third step of RNAi completes a circular pathway from the target to the initiator [2, 8]:
(iii) An aberrant mRNA resulting from step (ii) becomes a template for dsRNA produced by polymerization of RNA-dependent RNA polymerase (RdRp) (see the left semicircle of Figure 1).

Since each step is digital and circularly linked, RNAi resembles a kind of (digital) computation. This observation raises the question of whether RNAi can be viewed as a digital computation. If so, what is a computational meaning of RNAi and how computationally complex RNAi is. The purpose of this paper is to address these issues.

Firstly, we observe that RNAi can be modeled as a Minsky register machine. The Minsky register machine is a Turing complete model of computation, that (instead of an infinite tape for Turing machine) is equipped with two registers (for holding numbers) and a finite number of instructions (increment and decrement/jump) acting on the registers [10]. While most biological computational models to date are

Gabriel Ciobanu (Ed.): Membrane Computing and
Biologically Inspired Process Calculi (MeCBIC 2012)
EPTCS 100, 2012, pp. 107–112, doi:10.4204/EPTCS.100.8

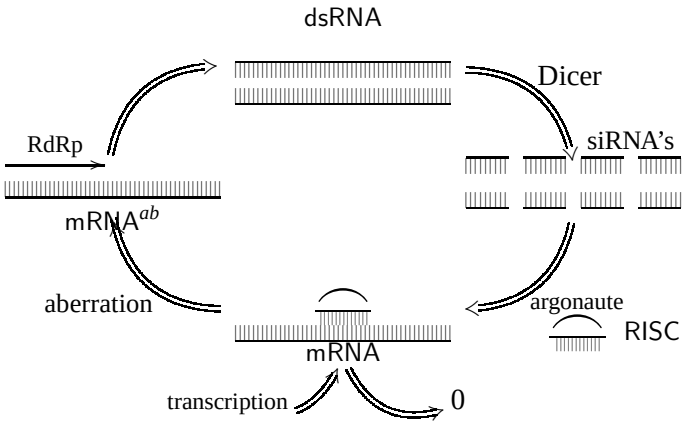

Figure 1: RNA interference

based on the Turing machine model; that is, they regard DNA as analogous to a single tape [7], the Minsky machine interpretation proposed here is intrinsic to the RNAi mechanism, whereby RNAs can be single or double stranded. We first present a naive machine model of RNAi, designated $\mathbf{RM}_{\mathsf{RNAi}}$, in which the two registers are realized respectively as the initiator (dsRNA) and the target (mRNA) of RNAi. Increment/Decrement instructions on the registers represent chemical reactions mediated by enzymes and proteins (e.g., RdRp and transcription/Dicer and RISC). However, the naive model lacks any rigorous computational language, hence requires a syntactical analysis. Capturing RNAi as a computational structure, such analysis aims to extract the computational meaning, in particular, the complexity, of RNAi.

Motivated by the work of Zavattaro-Cardelli [15], we describe our machine $\mathbf{RM}_{\mathsf{RNAi}}$ in the calculus of Chemical Ground Form (CGF), which is a minimal fragment of Milner's CCS equipped with interaction rates for each channel, and hence constitutes a subset of the stochastic π-calculus [11]. Introduced by Cardelli [3], CGF represents chemical kinetics by giving correspondence to a stochastic semantics of continuous time Markov chains. Despite its simplicity, the model sufficiently describes chemical kinetics compositionally. However, the primitive description of CGF lacks any direct representation of zero-tests for the registers, creating a tendency for the instructions of encoded $\mathbf{RM}_{\mathsf{RNAi}}$ to allow incorrect jumps. To avoid such erroneous probabilistic jumping, an inhibitor must be incorporated into the machine instructions. Biologically, this corresponds to a process known as *recursive* RNAi (recRNAi), an extension of RNAi [9, 12, 14], whereby siRNA produced and accumulating during RNAi inhibits not only mRNA but also RISC and Dicer. The extension to *recursive* RNAi (recRNAi) is obtained by adding a feedback linkage to RNAi. The recRNAi is directly represented by a register machine $\mathbf{RM}_{\mathsf{recRNAi}}$, in which siRNAs interactions are naturally interpreted as instruction inhibitors. We describe the machine in terms of CGF with fixed points. Probabilistic termination is then investigated in the recRNAi encoded system, and Turing completeness up to any degree of precision is demonstrated.

2 A Naive Interpretation of RNAi in Minsky Register Machine

In this section, we show that RNAi is naively interpreted as Minsky register machine [10].

Definition 2.1 (Register machine $\mathbf{RM}_{\mathsf{RNAi}}$ interpreting RNAi (cf. Figure 2)) RNAi is interpreted in the Minsky register machine $\mathbf{RM}_{\mathsf{RNAi}}$ as follows: Registers r_1 and r_2 hold species dsRNA and mRNA re-

spectively so that the increment on r_1 (res. r_2) produces one dsRNA (res. one mRNA) and the decrement on r_1 (res. r_2) removes one dsRNA (res.one mRNA). In biological terms, the increment on register r_1 represents *polymerization* RdRp with an aberrant mRNA template, while an increment on r_2 represents *transcription*. A decrement on r_1 models the *enzyme* Dicer which cleaves dsRNA into siRNAs, and a decrement on r_2 models the complementary *degradation* of mRNA by RISC. [1]

Figure 2: Register Machine $\mathbf{RM}_{\mathsf{RNAi}}$

The following table displays the chemical reactions for the increment/decrement on the two registers, where $\mathsf{mRNA}^\bullet$ denotes either 0 or mRNA^{ab}.

	r_1		r_2	
increment	(polymerization)	$\mathsf{RdRp} + \mathsf{mRNA}^{ab} \longrightarrow \mathsf{dsRNA}$	(transcription)	$\longrightarrow \mathsf{mRNA}$
decrement	(cleavage)	$\mathsf{dsRNA} + \mathsf{Dicer} \longrightarrow \mathsf{siRNA}'s$	(degradation)	$\mathsf{mRNA} + \mathsf{RISC} \longrightarrow \mathsf{mRNA}^\bullet + \mathsf{RISC}$

Table 1: chemical reactions

3 RNAi as Chemical Reaction and Register Machines

In this section, we describe the register machine $\mathbf{RM}_{\mathsf{RNAi}}$ in Section 2 in terms of of CGF. Recall that CGF is a subset of π-calculus and of CCS supplemented with channel transition rates. Using three interaction prefixes $\pi := \tau_{(r)}$, $?a_{(r)}$ and $!a_{(r)}$, CGF models collision between molecules as well as molecular decay. The parenthesized subscript (r) denotes the reaction rate of the channel. Collision and decay are described by

(decay of molecule) $\quad \cdots \oplus \tau_{(r)}.Q \oplus \cdots \quad \longrightarrow \quad Q$

(collision of molecules) $\quad \cdots \oplus ?a_{(r)}.Q \oplus \cdots \mid \cdots \oplus !a_{(r)}.R \oplus \cdots \quad \longrightarrow \quad Q \mid R$

Then a CGF is a pair (E,P) of a set E of *reagents* and a initial *solution* P. A reagent $X_i = M_i$ for naming a chemical specie and *molecules* M_i for describing the interaction capabilities of the corresponding species. Solution is a multiset of variables, which is released by interactions:

(Reagents) $E := 0$ and $X = M, E$ (Molecule) $M := 0$ and $\pi.P \oplus M$ (Solution) $P := 0$ and $X \mid P$

Formally, computation of CGF is defined in terms of Labelled Transition Graph, as defined in [3].

Every increment instruction $I_i = \mathsf{Inc}(r_j)$ is formalized directly for $j \in \{1,2\}$ so that once the chemical reactions of the first row of Table 1 are complete, we proceed to the next instruction I_{i+1}.

(Increment $I_i = \mathsf{Inc}(r_j)$)

$$I_i = \quad \mathsf{RdRp} \mid \tau.I_{i+1} \quad j = 1$$
$$I_i = \quad \mathsf{mRNA} \mid \tau.I_{i+1} \quad j = 2$$

[1] The machine interpretation assumes that the two species of dsRNA and mRNA are disconnected, so that the decrement and increment of either species induces no effect on the other. This assumption is justified because the synthesis of dsRNA is here regarded as primer-*independent* only [1, 2]; in other words, dsRNA is directly duplicated in the absence of primer. In primer-*dependent* dsRNA synthesis, the disconnection of the two species is violated. In such scenario, siRNA triggers polymerization, hence enables RdRp to copy a normal mRNA. See also the author's [6] on the difference of the two syntheses.

The decrement operations are more subtle. Decrements on on r_1 and on r_2 represent the chemical reactions of the second row of Table 1, which reactions ensure that Dicer and RISC interact to the entities in dsRNA and mRNA respectively, and thereby eliminate them. Although Dicer and RISC both induce decremental operations, RISC is recycled during degradation so that it is retained in the right-hand-side of (degradation), while the Dicer catalyst is consumed during the reaction (cleavage).

So that the registers may be decremented, they are interpreted as follows:

$$\begin{array}{ll} \text{Register } r_1 & \mathsf{dsRNA} \quad :=?a_1.(\mathsf{siRNA} \mid \cdots \mid \mathsf{siRNA}) \\ \text{Register } r_2 & \mathsf{mRNA} \quad :=?a_2.(\tau.0 \oplus \tau.\mathsf{mRNA}^{ab}) \end{array}$$

They represent that dsRNA and mRNA disappear by formation of siRNA, and by degradation or aberration, respectively.

If the chemical reaction occurs in the presence of dsRNA (res. mRNA), we proceed to the instruction I_{i+1}. Otherwise (i.e. if the reaction does not occur because dsRNA is absent (res. mRNA)), a jump is made to the instruction I_s. Thus in a primitive description of CGF, every decremental instruction $I_i = \mathsf{DecJump}(r_j, s)$ is described by

(Decrement instruction $I_i = \mathsf{DecJump}(r_j, s)$)

$$\begin{array}{llll} j=1 & I_i & =!a_1.(0 \mid I_{i+1}) \oplus \tau.I_s & \text{with Dicer} =!a_1.(0 \mid I_{i+1}) \\ j=2 & I_i & =!a_2.(\mathsf{RISC}|I_{i+1}) \oplus \tau.I_s & \text{with RISC} =!a_2.(\mathsf{RISC} \mid I_{i+1}) \end{array}$$

The above recursive definition of RISC for $j = 2$ corresponds to the recycling of RISC described in the degradation.

The decremental instructions so defined contain an error; that accidental jumps to I_s occur even if the register is non-empty (i.e. in the presence of channel $?a_j$). This error results from the absence of zero-test of the registers, a test which cannot be directly formulated in terms of CGF. Such an absence has been previously noted by Soloveichik et al.[13], in their studies of stochastic chemical reaction networks. Lack of zero-test is a main origin of Turing incompleteness of CGF [15], and motivated Cardelli and Zavattaro to develop their Biochemical Ground Form [4] as a minimalistic Turing complete extension of CGF.

4 Recursive RNAi and Probabilistic Termination

In this section, we model recursive RNAi in order to improve the defect described in Section 3, that the CGF machine interpretation $\mathbf{RM}_{\mathsf{RNAi}}$ allows non-feasible jumps. We extend the RNAi mechanism to a recursive RNAi (recRNAi), whose register machine $\mathbf{RM}_{\mathsf{recRNAi}}$ is described in terms of CGF + fixed points. This interpretation guarantees a probabilistic termination of the machine. Via this extended mechanism, siRNAs produced and accumulating during interference targets not only mRNA but also Dicer and RISC. A schematic of this situation is presented in Figure 3, in which the usual RNAi are displayed to the left, but siRNAs are produced by both Dicer and RISC (which simultaneously degrades mRNA). The right hand of Figure 3 includes inhibition arrows from siRNA to Dicer and RISC. The mechanism is recursive because the RISC complex containing siRNA is being degraded besides acting as a degrading agent. The recursiveness of RNAi prevents the decrement operators of Section 3 from taking erroneous jumps, since siRNAs accumulating throughout the RNAi cycle work as inhibitors of the decrement operators.

In recRNAi, the chemical reactions involved in Dicer and in RISC are not only those in the second row of Table 1 but also those in Table 2. The first row of Table 2 and (cleavage) represent reciprocal interactions on Dicer such that Dicer either makes dsRNA disappear by cleavage or Dicer is degraded by siRNA. Similar reciprocal interactions for RISC between the second row of Table 2 and (degradation).

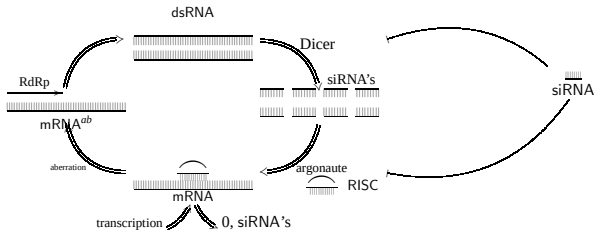

Figure 3: Recursive RNAi

(degradation of Dicer) $\mathsf{siRNA} + \mathsf{Dicer} \longrightarrow 0$
(degradation of RISC) $\mathsf{siRNA} + \mathsf{RISC} \longrightarrow 0$

Table 2: chemical reactions for recRNAi

We next configure recRNAi as a register machine $\mathbf{RM}_{\mathsf{recRNAi}}$ in terms of CGF with fixed points.

Definition 4.1 ($\mathbf{RM}_{\mathsf{recRNAi}}$ in CGF with fixed points)
Registers and $I_i = \mathsf{Inc}(r_j)$ are identical to those of Section 3. The decrement instruction, with incorporation of siRNA, is
(Decrement instruction $I_i = \mathsf{DecJump}(r_j, I_s)$)

$$I_i = !a_j.(0 \mid I_{i+1}) \oplus \tau.(!s.I_i \oplus \tau.I_s) = \mathsf{fix}_X.[a.(0 \mid I_{i+1}) \oplus \tau.(!s.X \oplus \tau.I_s)]$$
$$\mathsf{siRNA} = ?s.\mathsf{siRNA}$$

In the above definition of I_i, when $j = 1$ (res. $j = 2$), the left term $!a_j.(0 \mid I_{i+1})$ corresponds to Dicer (res. RISC) cleaving dsRNA (res. degrading mRNA), while the right term $\tau.(!s.I_i \oplus I_s)$ corresponds to Dicer (res. RISC) being degraded by siRNA. Hence our definition of I_i intrinsically reflects the reciprocal interactions of Dicer and RISC, and implies a recursive RNAi process in the presence of siRNA.
The fixed point definition of I_i derives from Zavattaro-Cardelli [15], but here we have highlighted a biological analogue of the definition. In the following, we modify slightly the results of [15] to obtain the main theorem of this section.

Given a state $(I_i, r_1 = l_1, r_2 = l_2)$ of register machine and a natural number h, the solution in $\mathbf{RM}_{\mathsf{recRNAi}}$ is defined by $[\![(I_i, r_1 = l_1, r_2 = l_2)]\!]_h := I_i \mid \prod_{l_1} \mathsf{dsRNA} \mid \prod_{l_2} \mathsf{mRNA} \mid \prod_h \mathsf{siRNA}$, where I_i on the right hand is that of Definition 4.1.

Proposition 4.2 (correspondence of computations between machine and $\mathbf{RM}_{\mathsf{recRNAi}}$) *Suppose a one step computation of register machine is given by $(I_i, r_1 = l_1, r_2 = l_2) \longmapsto (I_j, r_1 = l_1', r_2 = l_2')$. Then we have the following for the solutions of the two states of the computation:*

- *If $I_i = \mathsf{Inc}(r_j)$ or $I_i = \mathsf{DecJump}(r_j, s)$ with $l_j = 0$, then the solution $[\![(I_i, r_1 = l_1, r_2 = l_2)]\!]_h$ can converge to the solution $[\![(I_j, r_1 = l_1', r_2 = l_2')]\!]_h^{\dagger}$ with the probability 1.*
- *If $l_j > 0$ and $I_i = \mathsf{DecJump}(r_j, s)$, the solution $[\![(I_i, r_1 = l_1, r_2 = l_2)]\!]_h$ can reach to a solution $[\![(I_j, r_1 = l_1', r_2 = l_2')]\!]_k^{\dagger}$ for some natural number $k \geq h+1$ with the probability $> 1 - \frac{1}{h}$.*

Proof.
We illustrate the case of $I_i = \mathsf{DecJump}(r_j, s)$ (direct for increment instructions), where sigma in the second column denote the probability that $\mathbf{RM}_{\mathsf{recRNAi}}$ computations attain the right hand side solutions. The schematic in the third column displays the execution paths for the probability.

$l_j = 0$	$\sum_{i=0}^{\infty} (\frac{h}{h+1})^i \times \frac{1}{(h+1)} \quad = 1$	$I_i \underset{h}{\overset{1}{\rightleftarrows}} \bullet \xrightarrow{1} I_s = I_j$
$l_j \neq 0$	$\sum_{i=0}^{\infty} (\frac{1}{l_j+1} \times \frac{h}{h+1})^i \times \frac{l_j}{l_j+1} \quad > 1 - \frac{1}{h}$	$I_i \underset{h}{\overset{1}{\rightleftarrows}} \bullet \xrightarrow{1} I_s$, $\bullet \xrightarrow{l_j} I_{i+1} = I_j$

We now state the main theorem of this section.

Theorem 4.3 (probabilistic termination) *The following are equivalent:*

- *A Minsky register machine starting from a state* $(I_j, r_1 = l_1, r_2 = l_2)$ *terminates.*
- *A CGF* $(\mathbf{RM}_{\mathsf{recRNAi}}, [\![(I_j, r_1 = l_1, r_2 = l_2)]\!]_h)$ *probabilistically terminates with probability greater than* $1 - \sum_{k=h}^{\infty} \frac{1}{k}$.

Proof. Note first that following the execution of a decrement instruction, the number of siRNA increases by at lease one. This is because at least one siRNA is produced by Dicer cleavage or by RISC (as it degrades mRNA). By Proposition 4.2 a computation of register machine containing d decrement instructions is faithfully reproduced with probability greater than the following: $(1-\frac{1}{h})(1-\frac{1}{h+k_1})\cdots(1-\frac{1}{h+k_1+\cdots+k_d}) \geq \prod_{k=h}^{h+d}(1-\frac{1}{k}) > 1-\sum_{k=h}^{h+d}\frac{1}{k}$, where $k_i \geq 1$ is the number of siRNAs produced by the corresponding decrement instruction. □

References

[1] David Baulcombe, *RNA Silencing in Plants*, Nature. **431**, 356-63, (2004), doi:10.1038/nature02874

[2] Peter Brodersen and Olivier Voinnet, *The Diversity of RNA Silencing Pathways in Plants*, TRENDS in Genetics 22(5), 268280 (2006), doi:10.1016/j.tig.2006.03.003

[3] Luca Cardelli, *On Process Rate Semantics*, Theor. Comput. Sci. **391**(3): 190-215 (2008), doi:10.1016/j.tcs.2007.11.012

[4] Luca Cardelli and Gianluigi Zavattaro, *Turing Universality of the Biochemical Ground Form*. Mathematical Structures in Computer Science **20**(1): 45-73 (2010), doi:10.1017/S0960129509990259

[5] A. Fire, S. Xu, M. Montgomery, S. Kostas, S.Driver and C. Mello, *Potent and Specific Genetic Interference by Double-stranded RNA in Caenorhabditis Elegans*, Nature **391** (6669): 806811 (1998), doi:10.1038/35888

[6] Masahiro Hamano, "Sustainability of RNA Interference in Rule Based Modelling", To appear in Proc. of the third International Workshop on Static Analysis and Systems Biology (SASB 2012).

[7] Z. Ignatova, K-H. Zimmermann, I. Martínez-Pérez, "DNA Computing Models", Springer (2008)

[8] Richard A. Jorgensen, *RNA Traffics Information Systemically in Plants*, Proc. Natl. Acad. Sci. USA, **99**(18) 11561-11563 (2002), doi:10.1073/pnas.192449099

[9] W. F. Marshall, *Modeling Recursive RNA Interference*, PLoS Computational Biology, **4** (9) e1000183. (2008), doi:10.1371/journal.pcbi.1000183

[10] Marvin Minsky, "Computation Finite and Infinite Machines (1st ed.)." Englewood Cliffs, N. J.: Prentice-Hall (1967)

[11] C. Priami, A. Regev, E. Shapiro and W. Silverman, *Application of a Stochastic Name-passing Calculus to Representation and Simulation of Molecular Processes*, Information Processing Letters **80**, 25-31. (2001), doi:10.1016/S0020-0190(01)00214-9

[12] MW. Rhoades, BJ. Reinhart, LP. Lim, CB. Burge, B. Bartel, et al. *Prediction of Plant microRNA Targets*. Cell **110** 51320. (2002), doi:10.1016/S0092-8674(02)00863-2

[13] David Soloveichik, Matt Cook, Erik Winfree and Shuki Bruck, *Computation with Finite Stochastic Chemical Reaction Networks*. Natural Computing, **7** (4), 615-633. (2008), doi:10.1007/s11047-008-9067-y

[14] Z. Xie, KD. Kasschau and JC. Carrington, *Negative Feedback Regulation of Dicer-like1 in Arabidopsis by microRNA-guided mRNA Degradation*. Curr Biol **13**: 784789. (2003), doi:10.1016/S0960-9822(03)00281-1

[15] Gianluigi Zavattaro and Luca Cardelli, *Termination Problems in Chemical Kinetics*. Proc. CONCUR 2008, LNCS **5201**, Springer-Verlag. 477-491 (2008), doi:10.1007/978-3-540-85361-9_37